

City-Eyes: A Multi-source Data Integration Based Smart City Analysis System

Zijian Ma, Dan Lu, Qian Liu, Jingyuan Wang, Zhang Xiong

School of Computer Science and Information Technology, Beijing Jiaotong University, Beijing, 100044, China

School of Computer Science and Engineering, Beihang University, Beijing, 100191, China

Abstract—The project named after City-Eyes is based on the HBASE big data analyzing platform and the real-time-display technology with H5/CSS3, which could analyze data and result dynamically. The project is devoted to designing the online city-data-analyzing platform, with interacting in real time, filtering, analyzing and displaying data. In the view of filtering, we sample Wuxi and collect much information, including phone signaling, railway, airports, Weibo, population mobility and etc. Moreover, we write enough data mining scripts for varieties of data to extract the value of them. In the view of analyzing, we build an efficient analyzing model, which divides the task into two parts, which includes the front end and the back end. The back end is focusing on compressing data and extracting valid information by clustering the data streams. At the same time, the front-end applies the real-time-display technology to provide an interactive and nice-looking data visualizing solution for our users [1]. In the view of displaying, we come up with various solutions from two angles for every variety of data: 'distribution and network'. With svg/canvas animation tech, the data can be mixed into the map and display itself dynamically [2].

I. INTRODUCTION

City-Eyes system serves for urban construction researchers. It provides a powerful, friendly and simple operated web platform. There are total three modules in this system, each of them is aimed at a hot city issue or research theme. They are City-Graph, Traffic-Network and Crowd-Tracker. To provide the service of GIS with data display, we bind the map with Baidu's open API. As a result, locating POI filtering and analyzing could appear in the data map [3]. We also provide an usage video in home page to help new users understand it. Here is the detail introduction of each module.

II. CITY-GRAPH

City-Graph uses a heat map to show the distribution of city population during the time evolution, which can help us analyze the abnormal crowd mass in our city. We also use Weibo data to draw a word cloud graph to excavate key factors behind each mass event [4]. The main data source of this module are phone data, Weibo data and static GIS data. Here are some details of its functions.

A. Population distribution heat map

In this function, we can find out the active phone records distribution in the whole city, so that we can know where the busiest area in the city is. Each area in the middle map stands for a TAZ (Traffic Analysis Zone, which divide by government

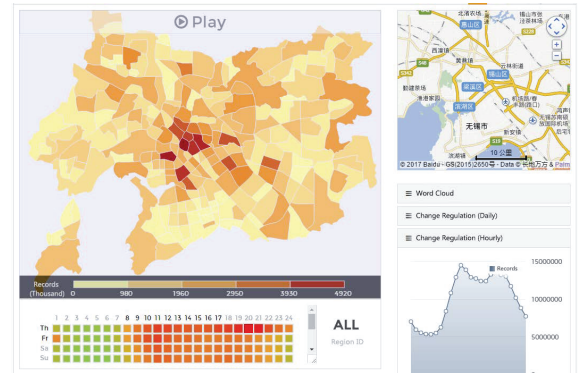


Fig. 1: Distribution - Population heat map

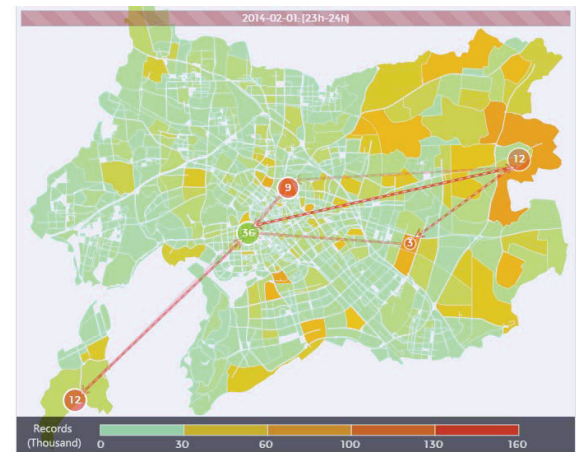


Fig. 2: Route - Gathering place evolution

benchmark). Each box in the bottom graph stands for an hour. The right map is a linked GIS map powered by Baidu, when you select one area in the middle map, you can get details information in it. Also, we provide a line chart to show the evolution of active phone records during one day.

As Fig. 1 shows, there is a 'Play' button over the middle map. When we click it we can view the gathering place evolution. It shows the migration of busiest area in this city dynamically. We can easily find some crowd incidents from this function.

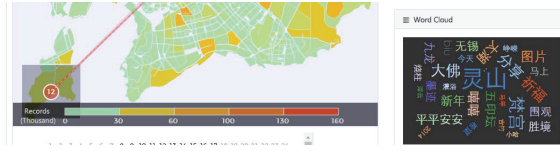


Fig. 3: Graph - Word cloud

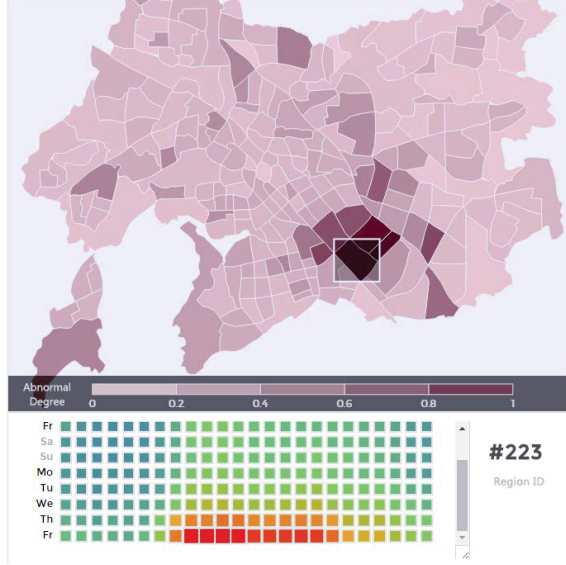


Fig. 4: Distribution - Abnormal area detecting map

B. Word cloud graph

When we select an area, this function could display some hot words that people there chat most frequently. We draw it as a word cloud. In this case, we can conclude that a New Year countdown party which associated with Lingshan Grand Buddha held here, which caused this crowd incident.

C. Abnormal area detecting

In this function, with deep learning methods, we can detect abnormal areas in this city which the population records here didn't obey the regular pattern in most areas during the selected time span. In this case, we select an detected area (mark with deep color) and find the records here suddenly reached peak in Friday. In fact this day is the first day after Chinese New Year holidays, and there is airport in this area, many people worked here come back by plane in this day so that records here changed irregularly.

III. TRAFFIC-NETWORK

Traffic-Network uses a dynamic network OD graph to show the traffic status in our city. In addition, we overlay taxi data, phone data and dangerous goods transport data to find dangerous area in the city. The main data source of this module are phone data, taxi (floating cars) data, dangerous goods transport data. Here are some details of its functions.

A. Dynamic traffic OD map

In this function, we provide a dynamic map to show the traffic condition in the city. We can display the whole OD

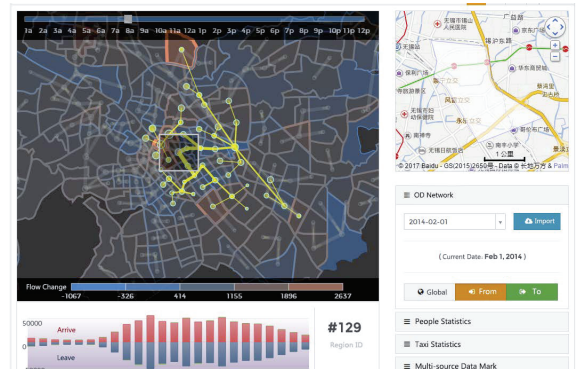


Fig. 5: Network - Traffic OD map

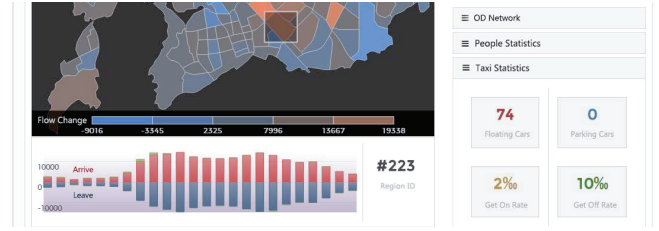


Fig. 6: Taxi statistics

(origin and destination path) network to figure out which area has hard traffic pressure. In addition, we can show where people come from (by yellow lines) and where they go (by green lines) when we select an area in 'Come from' module or 'Go to' module.

B. People and Taxi statistics

When we select one area, we can see how many persons arrived/leave here during 24 hours and some taxi statistics, such as the count of floating cars, the count of static cars, and the last two proportions show that how many persons get on or get off taxi in 1,000 persons. This will be helpful when researching the taxi service status [5].

C. Dangerous area detecting

Overlapping dangerous goods transport data and floating cars data, we can easily find out some dangerous areas. In this case, the area signed with circle has too many social cars (marked by a red hexagon) and dangerous goods transport vehicles (marked by a big exclamation mark icon). It's a symbol of traffic dangers, which could lead to safety accidents [6]. As we all know, some places that trucks unload are too close to residential area. This is also a constant safety issue; we can find it effectively by using phone data and dangerous goods transport data in same way.

IV. CROWD-TRACKER

Crowd-Tracker can deeply display the crowd behavior by analyzing there phone data in some time. It draws all the Business-Home path of each person, which can show the distribution of working area and living quarters. It can also classify the crowd from different angles, and display the



Fig. 7: Distribution - Dangerous area detecting map



Fig. 8: Network - Business-Home path

proportion by pie chart. The main data source of this module are phone data (long term), POI data. Here are some details of its functions.

A. Business-Home path based network graph

When we analyze one person's phone records in a long term, we can conclude his working place and living place. Starting from this idea, we analyzed a lot of phone records and drew all the paths from their working places (yellow point) to living places (blue point) in one map. Moreover, this can help us find out where most of people live or work in this city [7]. We also provide a direct view (using icon mark) to show the distribution of living areas and working areas as Fig. 9.

B. People classification in each area

Each person has its own travel mode. Someone travel between working place and living place, someone always stay at home, and someone like couriers and drivers travel among many places [8]. This mode can be reflected in their phone records. By analyzing these records we can figure out the proportion of each mode in different areas. Exactly, we could even know whether one person is local or lot by analyzing his phone records during Chinese New Year holidays (strangers would go back to their hometown and their records lost). We show it in our system interface.

V. CONCLUSION

In this paper, we make a simple introduction for City-Eyes system. Moreover, we proposed a new method for smart



Fig. 9: Distribution - Home/Business Area



Fig. 10: Pie chart - People classification

city analysis, that's multi-source data integration. Big data visualization and deep learning provide us a solution for urban planning and other safety issues, so we powered this web platform to realize it. As a result, this system detected many potential problems for our sample city - Wuxi. Not only we could find out the reason of some crowd accidents, but we could also detect dangerous areas and research people travel modes. Geography information data, travelling records (both phone data and vehicles GPS data) and social media chatter help us draw a living picture of our city, we could clearly see its health condition and evolution in this way.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (Grant No.61572059, 61202426), the State Key Program of National Natural Science of China (Grant No. 71531001), and the Basic Science Research Program of Shenzhen (Grant No. JCYJ20150624154400509).

REFERENCES

- [1] T. Lehtipalo, J. Thoreson, T. Fortes, and D. Vulcan, "Multi-source data visualization system," 2012.
- [2] J. Zhou, Y. Chen, X. Zhang, and W. Zhan, "Modelling the diurnal variations of urban heat islands with multi-source satellite data," *International Journal of Remote Sensing*, vol. 34, no. 21, pp. 7568–7588, 2013.
- [3] Y. Zheng, "Introduction to urban computing," *Geomatics and Information Science of Wuhan University*, vol. 40, no. 1, pp. 1–13, 2015.
- [4] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *Acm Transactions on Intelligent Systems and Technology*, vol. 5, no. 3, p. 38, 2014.
- [5] D. Ashbrook and T. Starner, "Using gps to learn significant locations and predict movement across multiple users," *Personal and Ubiquitous Computing*, vol. 7, no. 5, pp. 275–286, 2003.
- [6] M. A. Bayir, M. Demirbas, and N. Eagle, "Discovering spatiotemporal mobility profiles of cellphone users," in *World of Wireless, Mobile and Multimedia Networks and Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*, 2009, pp. 1–9.
- [7] M. J. Bradley S, "Comparison of energy use and co2 emissions from different transportation modes," 2007.
- [8] K. Dufkov, J. Y. L. Boudec, L. Kencl, and M. Bjelica, *Predicting User-Cell Association in Cellular Networks from Tracked Data*. Springer Berlin Heidelberg, 2009.